IT Unit 3

Topic 2

# Ch 2, Data Analytics: Drawing Conclusions, Part 1

**Integrity of data,** p 111- 121

1. Why is data integrity important? Explain possible effects of using data that lacks integrity.

* To produce useful, accurate, timely and relevant data the data we are using must have integrity. If the data isn’t authentic or lacks integrity we might of false, old and simply irrelevant data which will make our reports ineffective.

**Timeliness**

1. What does timeliness refer to? Give some examples of data that is not timely.

* It means data has been processed while it is current, and there are no significant delays in receiving it. Examples of data that isn’t timely would be doing a report on the Sydney Swans afl list and using their senior list from the year 2005.

**Authenticity**

1. What are the characteristics of authentic data?

* Comes from the author and/or source it claims to be from
* Has not been deliberately corrupted
* Is not fake or disguised as something else
* Has not been changed since authorisation

1. List some challenges to the authenticity of data.

* Spoofing
* Plagiarism
* Viral ads and fake YouTube videos
* Computer generated imagery

1. Authenticity techniques. How can you authenticate data, both digital and non-digital

**Digital**-

* Use SSL and TLS protocols
* Use checksums
* Use email validation

**Non-digital**-

* Contact original authors
* Compare original documents with allegedly accurate reports

**Relevance**

1. What makes data relevant?

* Relevance is how closely a resource, such as a book, database or webpage, corresponds to the people’s desire for information.

**Accuracy**

1. Distinguish between content and form in terms of accuracy of data.

* Content refers to functionality and how well the solution works and whether it has met the success criteria, while form refers to appearance and how well the solution has been laid out and whether it is easy to navigate through and easy to look at.

1. Briefly elaborate on the following challenges to data accuracy
   1. Correctness

* Refers to the values that are stored on the solution and that they must be correct.
  1. Completeness
* Completeness means that your data set is just that: complete. It means you have all data from all research participants on all variables at all relevant points in time and space.
  1. Clarity
* Clarity is about formatting the data in an unambiguous manner to prevent misinterpretation.
  1. Consistency
* Correct, unambiguous data can still cause a problem to the database if it is not consistent. Inconsistent data is unwelcome because it means the data is unreliable. This is why consistency needs to be used throughout the database and/or solution.

1. Measures to improve accuracy, briefly elaborate on the following measures;
   1. Correctness

To improve correctness, you can perform data quality assurance to cleanse or scrub data. This will identify and remove or repair data that is incomplete, inaccurate, irrelevant or inconsistent, it may also standardise data. E.g. Street-St.

* 1. Completeness
* For data that you suspect has gaps because it has been collected but no published in full, you could contact the original data collectors to ask if they have unpublished data that they could provide you.
  1. Clarity
* A method to improve the clarity of the database is to enforce data formatting and validation rules in your spreadsheets and database that prevent misinterpretation of the data. Try to avoid rekeying the data where possible – copy and paste or import instead.
  1. Consistency
* To prevent inconsistency in your solution, formulate the same question that means the same thing in different ways to check the respondents consistency when answering a question, and also enforce consistent data formatting and validation rules.